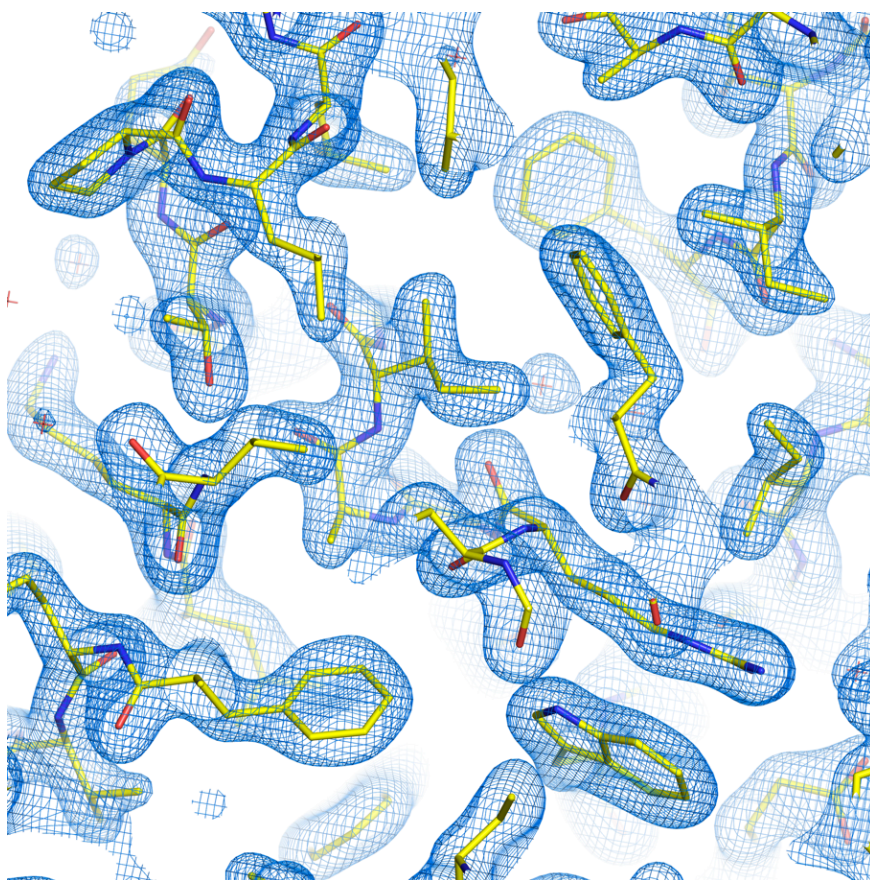
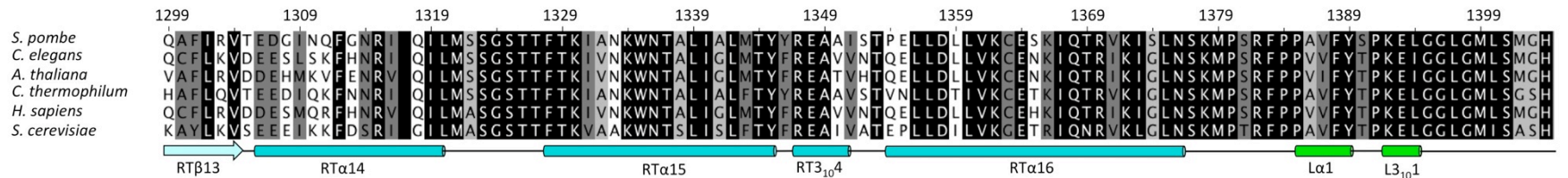
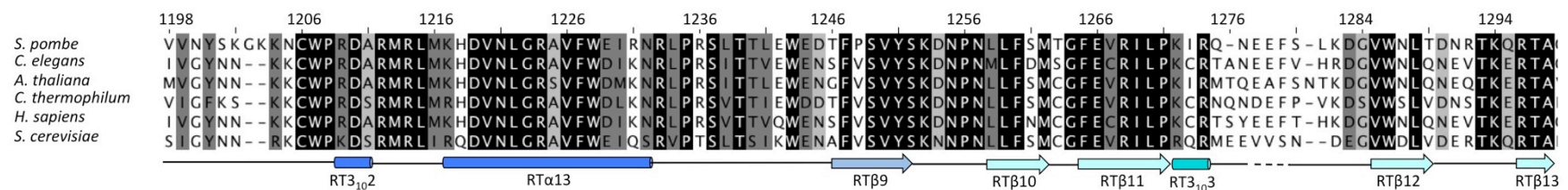
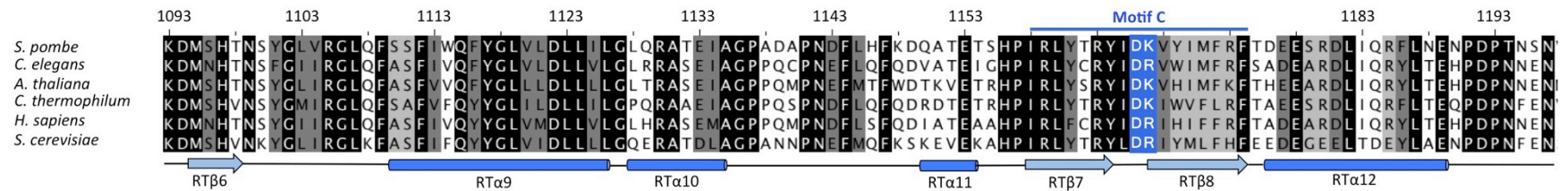
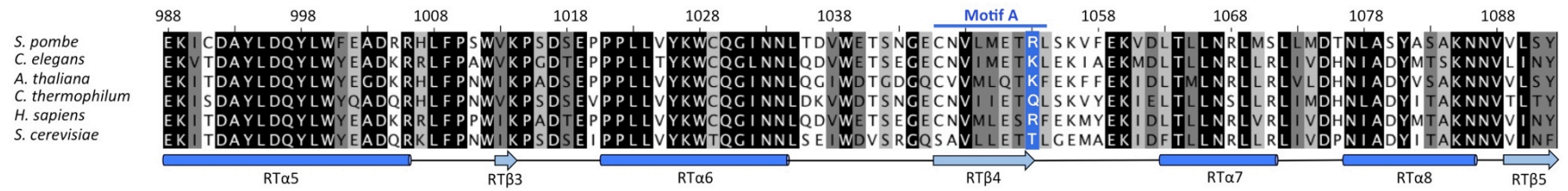
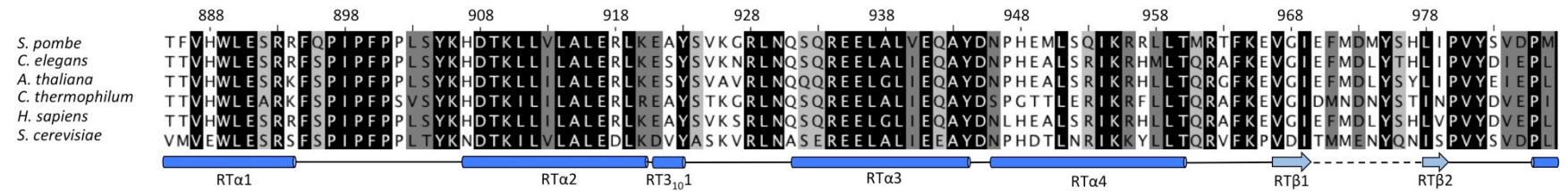
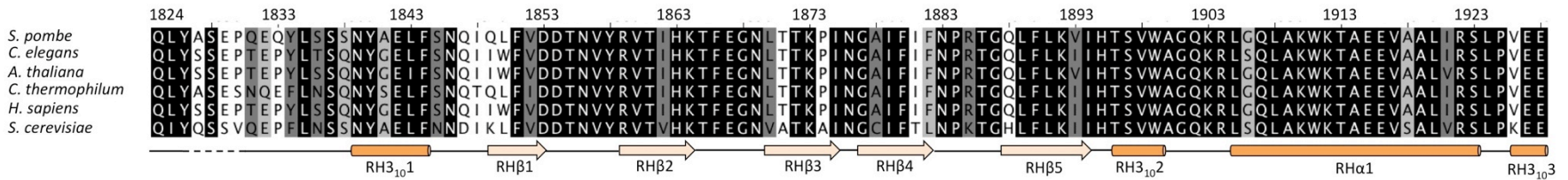
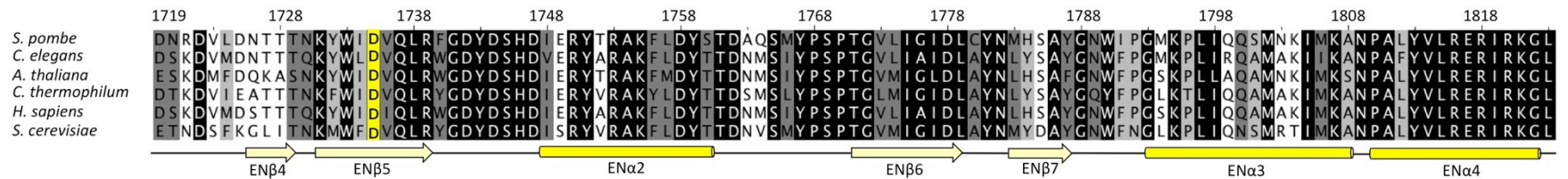
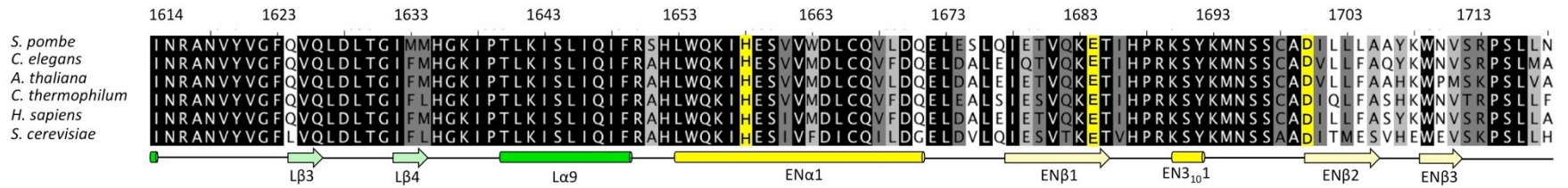
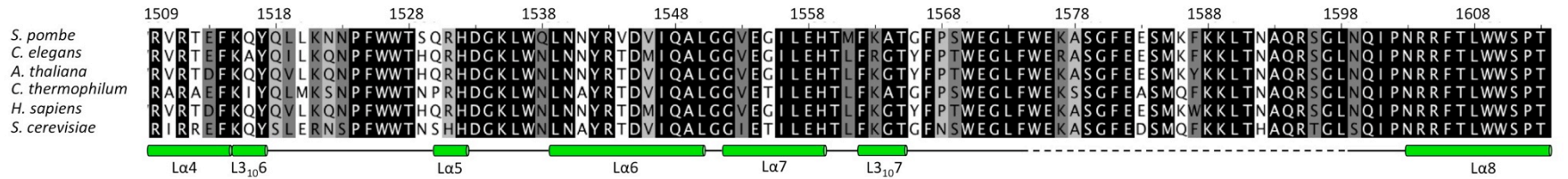
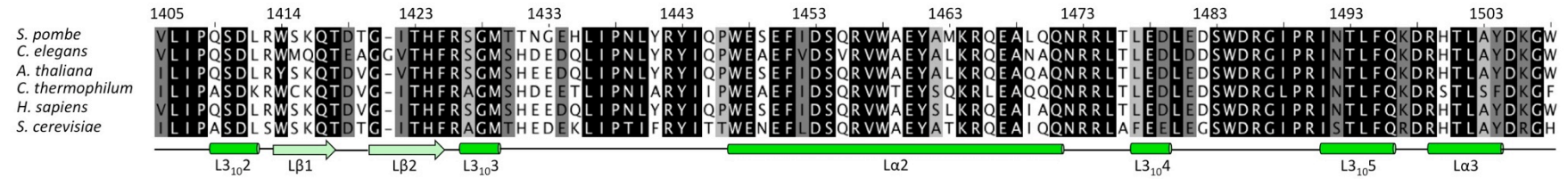


**Figure S1 | Validation of the molecular replacement solution.** Anomalous electron density from methyl mercury derivative crystal (contoured at  $10\sigma$ ) superimposed on the MR solution. All anomalous peaks are located in the close proximity to the cysteine residues in both molecules of the  $P2_12_12_1$  asymmetric unit: **a,b**, C62 in Aar2; **c,d**, C1878 in the Prp8 RNaseH-like domain; **e,f**, C2288 in the Prp8 Jab1/MPN domain.

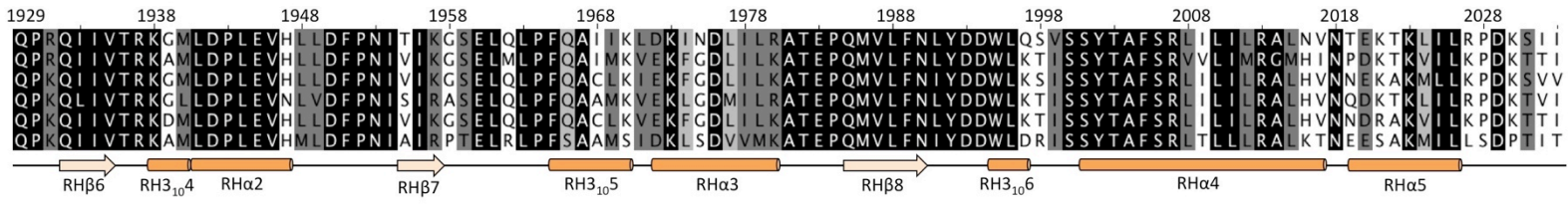


**Figure S2 | The 2Å resolution electron density map of the Prp8<sup>885-2413</sup>-Aar2 complex in the C222<sub>1</sub> space group. The 2Fo-Fc map is contoured at 1.5  $\sigma$ .**

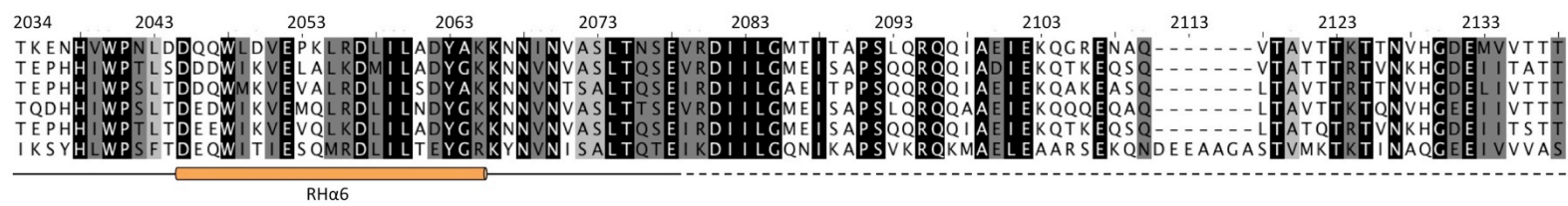




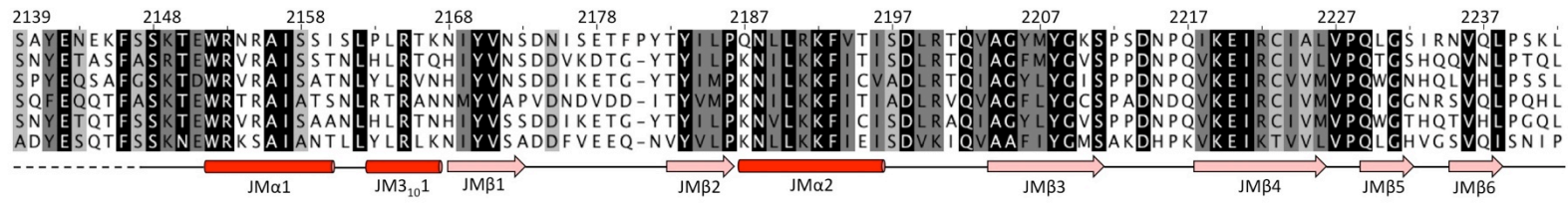
*S. pombe*  
*C. elegans*  
*A. thaliana*  
*C. thermophilum*  
*H. sapiens*  
*S. cerevisiae*



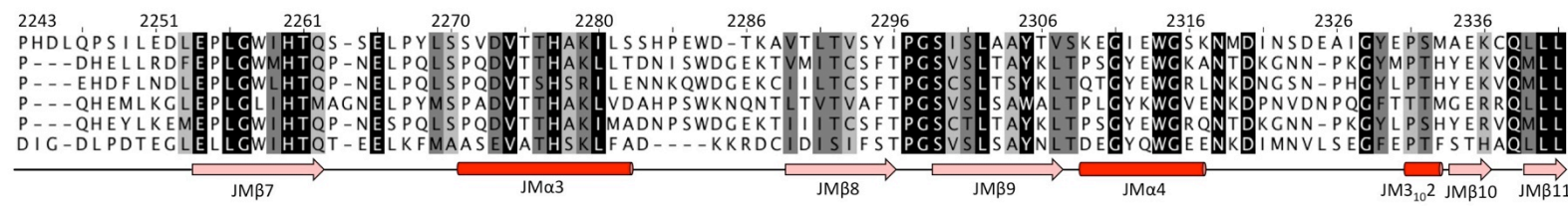
*S. pombe*  
*C. elegans*  
*A. thaliana*  
*C. thermophilum*  
*H. sapiens*  
*S. cerevisiae*



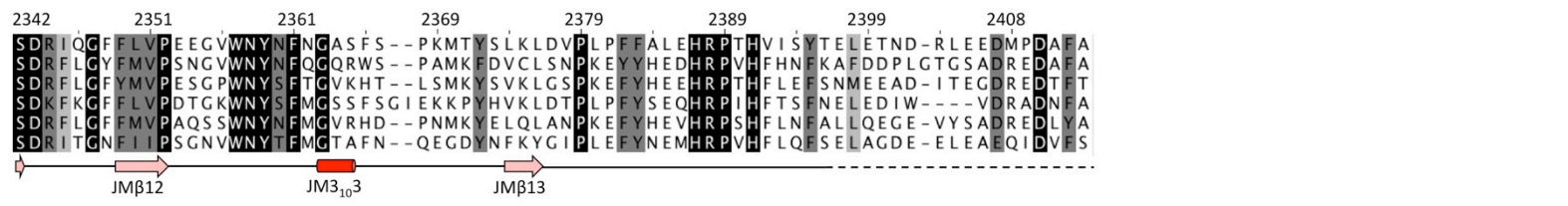
*S. pombe*  
*C. elegans*  
*A. thaliana*  
*C. thermophilum*  
*H. sapiens*  
*S. cerevisiae*



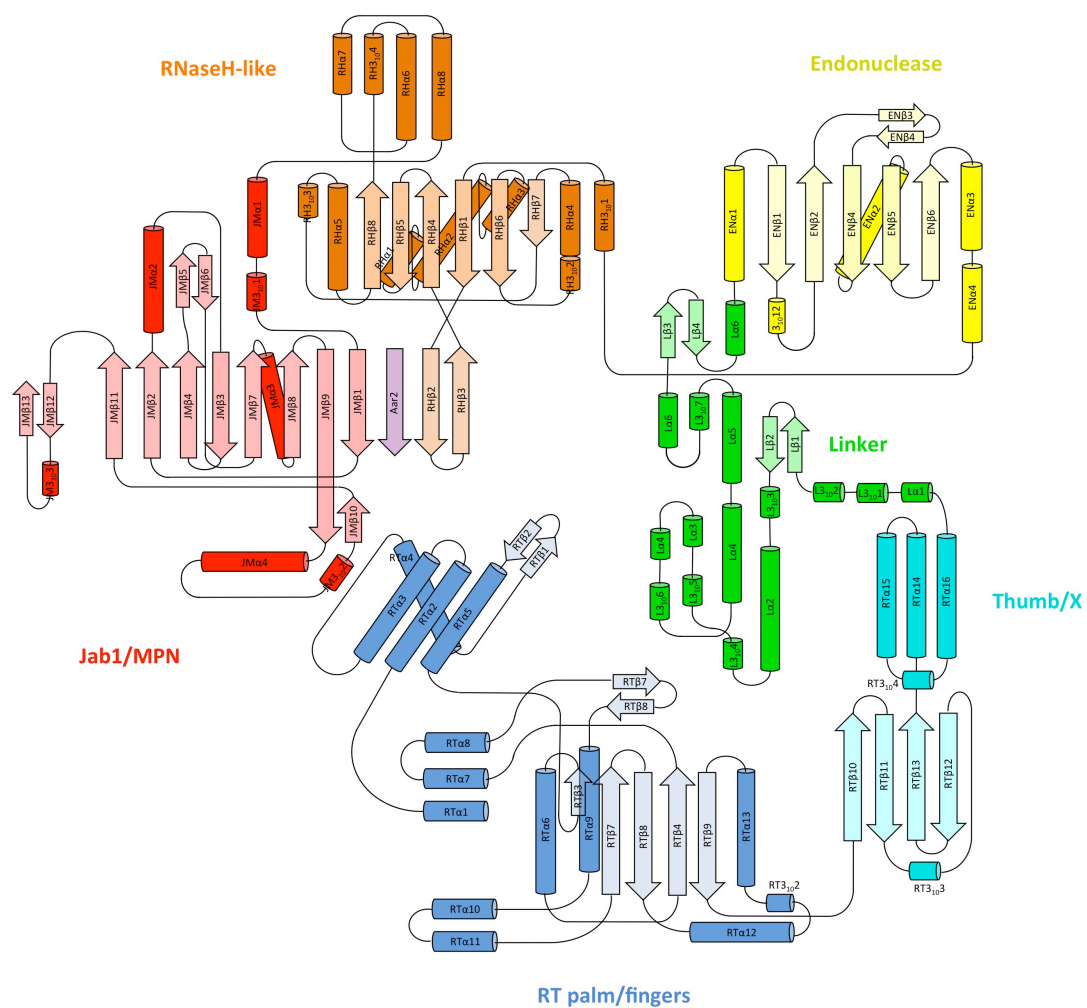
*S. pombe*  
*C. elegans*  
*A. thaliana*  
*C. thermophilum*  
*H. sapiens*  
*S. cerevisiae*



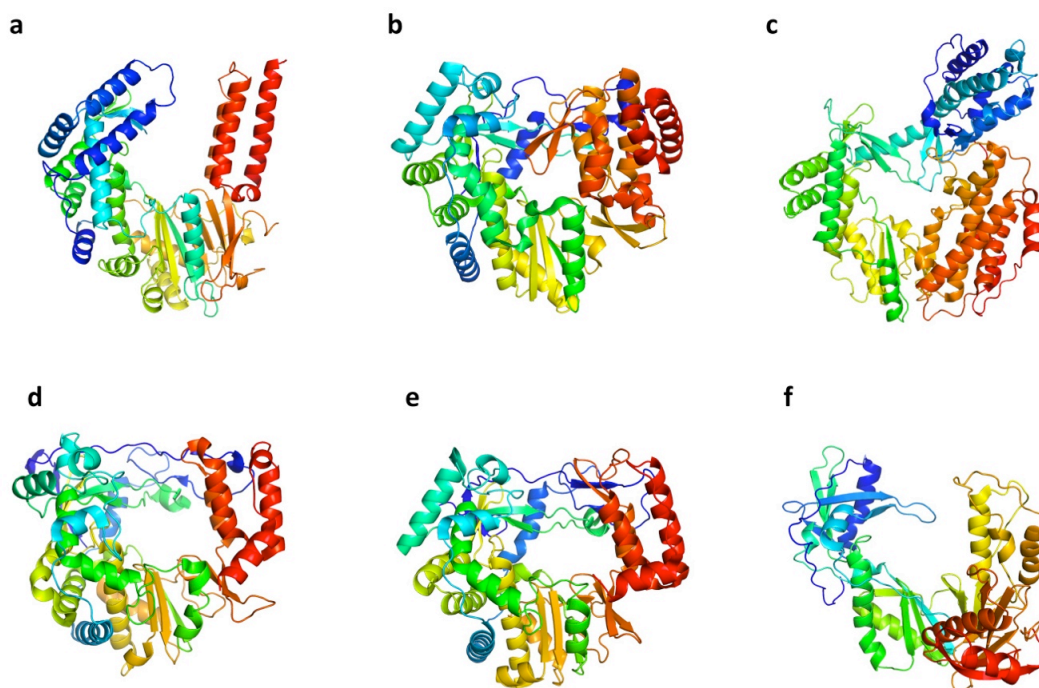
*S. pombe*  
*C. elegans*  
*A. thaliana*  
*C. thermophilum*  
*H. sapiens*  
*S. cerevisiae*



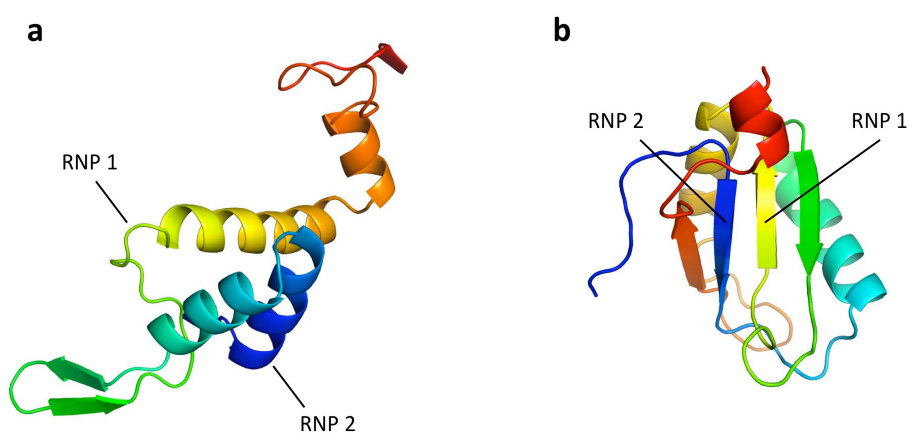
**Figure S3 | Multiple sequence alignment of Prp8 from various organisms.** Secondary structure elements are represented by blocks and arrows, with those in each domain coloured and numbered separately and given prefixes: RT, reverse transcriptase domain (blue); L, linker domain (green); En, endonuclease domain (yellow); RH, RNaseH-like domain (Orange); JM, Jab1/MPN domain (red). Dashed lines represent regions with missing electron density. Putative catalytic residues in the RT and En domains are highlighted in blue and yellow, respectively.



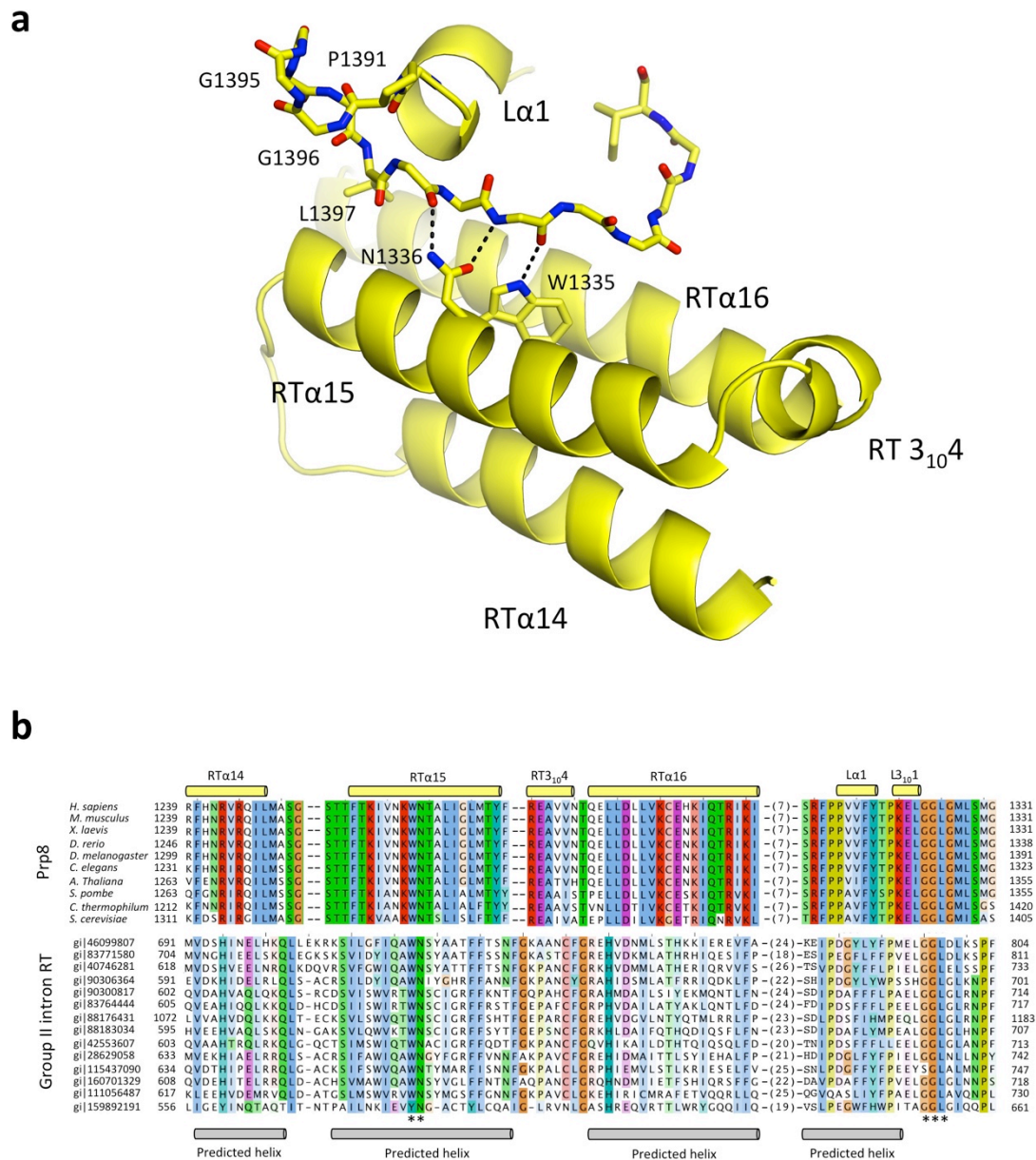
**Figure S4 | Topology plot for Prp8 residues 885-2413.** Secondary structure elements are coloured and numbered as in Figure S3.



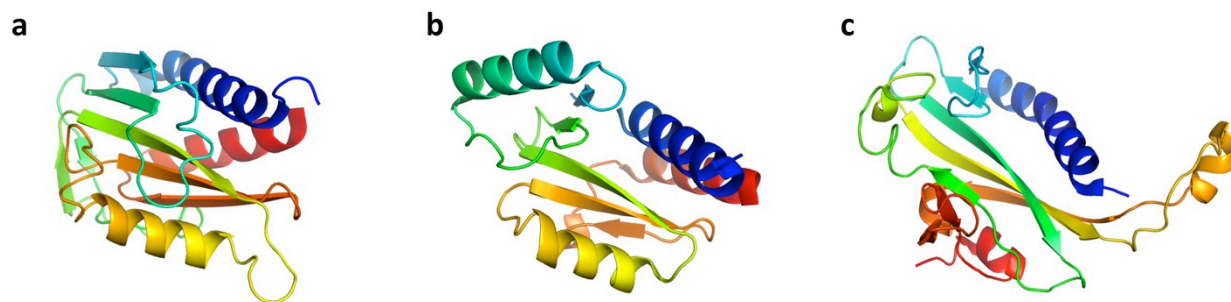
**Figure S5 | Comparison of the Prp8 RT domain and various polymerases.** **a**, Prp8<sup>885-1375</sup>; **b**, HCV (Hepatitis C virus) NS5B RNA-dependent RNA polymerase (3SKE); **c**, *Tribolium castaneum* telomerase catalytic subunit TERT (3DU5); **d**, Polio Virus 1 RNA dependent RNA polymerase (3OL8); **e**, Norovirus RNA dependent RNA polymerase (3NAH); **f**, HIV-1 Reverse Transcriptase (3HVT).



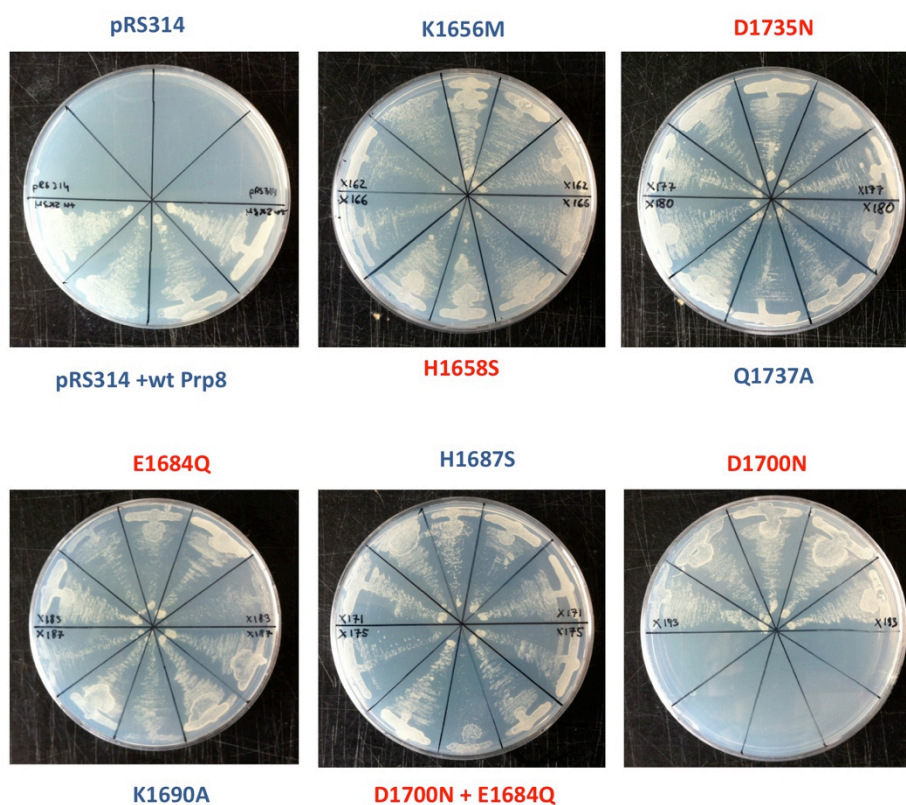
**Figure S6 | Predicted fold in Prp8 sub-domain.** **a** fragment of Prp8 (residues 1059-1151) predicted to form RNA Recognition Motif (RRM)<sup>21</sup>; **b**, canonical RRM as represented by U1A protein (1URN).



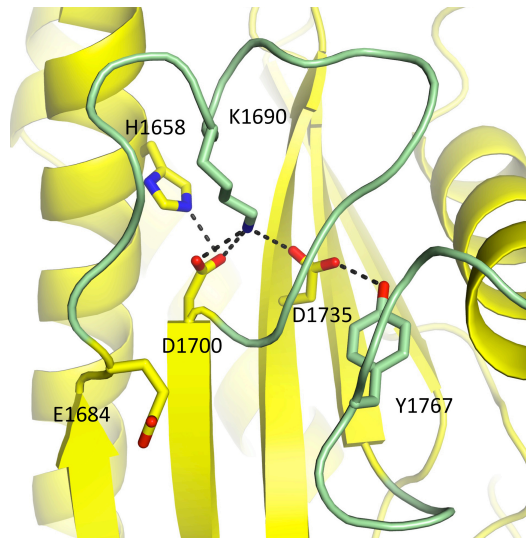
**Figure S7 | The structure of the Th/X domain of Prp8. a**, The helix bundle in the Th/X sub-domain of Prp8. The residues of Prp8 that play crucial structural roles are conserved in the reverse transcriptase Th/X domain of fungal group II intron-encoded protein. The Trp1335 and Asn1336 side chains form hydrogen bonds with the main chain amide and carboxyl groups to tether a long extended polypeptide chain (only main chain atoms are shown for clarity) which extends into the active site cavity of Prp8. Pro1382 and Pro1391 flank the short helix (La1). Gly1395 and Gly1396 allow a sharp turn of the main chain and Leu1397 interacts with hydrophobic residues on the helices. **b**, Sequence similarity between the Th/X sub-domain of Prp8 and group II intron encoded protein. Adapted from Dlakic and Mushegian<sup>27</sup>.



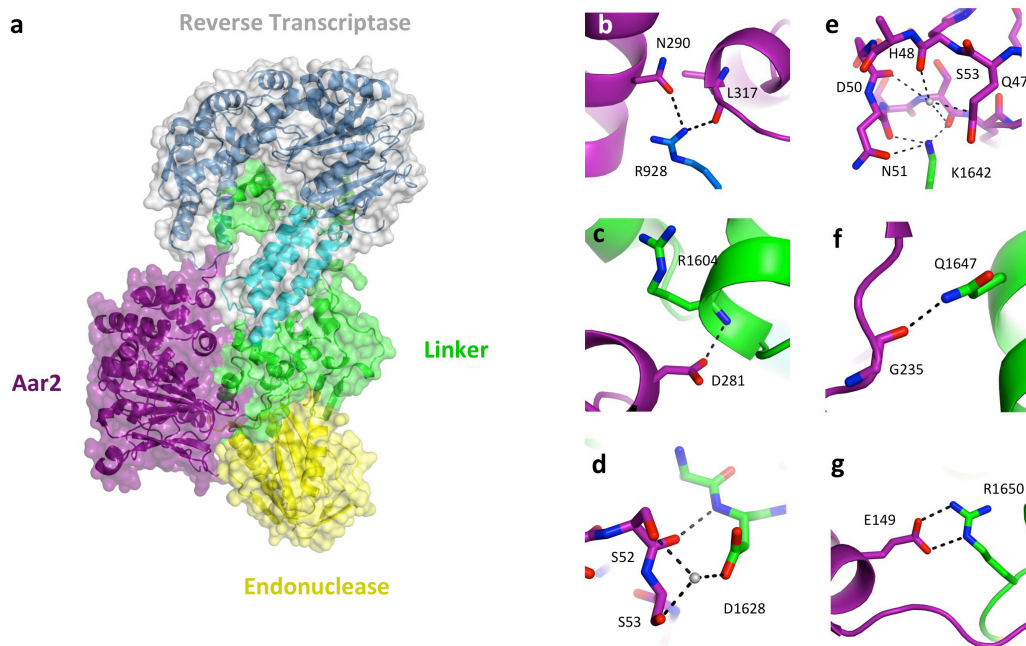
**Figure S8 | Type II restriction endonuclease fold.** **a**, Prp8 Endonuclease domain (1650-1810); **b**, Influenza virus PA endonuclease (3HW4, chain B, residues 30-179); **c**, restriction endonuclease Eco RV (1RVE, chain A, residues 35-205)



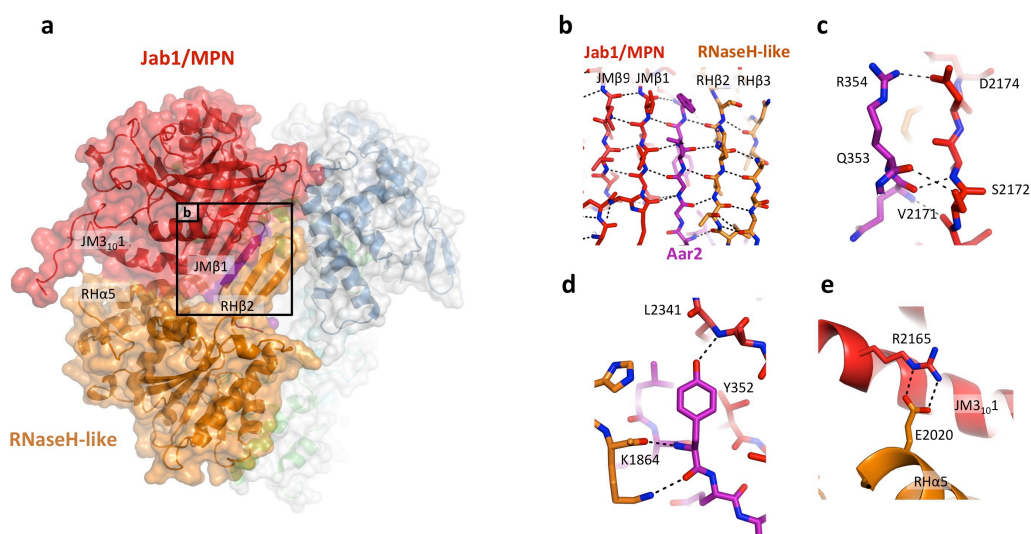
**Figure S9 | Mutagenesis of the catalytic residues (red) in Prp8 En domain.** The effects of each mutation were analysed in 5 replicates by the plasmid shuffling method. No significant effect on viability was observed.



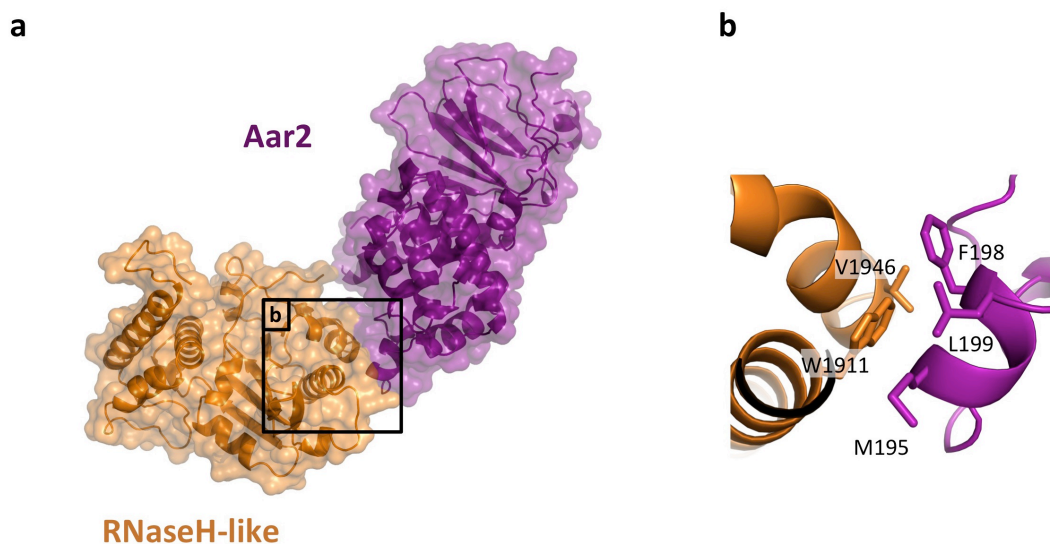
**Figure S10 | Polypeptide loops block the Prp8 Endonuclease active site.** Putative catalytic residues are involved in the network of polar interactions to stabilize the long polypeptide loops coloured in green (residues 1685-1699 and 1762-1771). These loops exhibit extraordinary sequence conservation, suggesting their functional importance. The “catalytic residues” may be conserved for structural reasons.



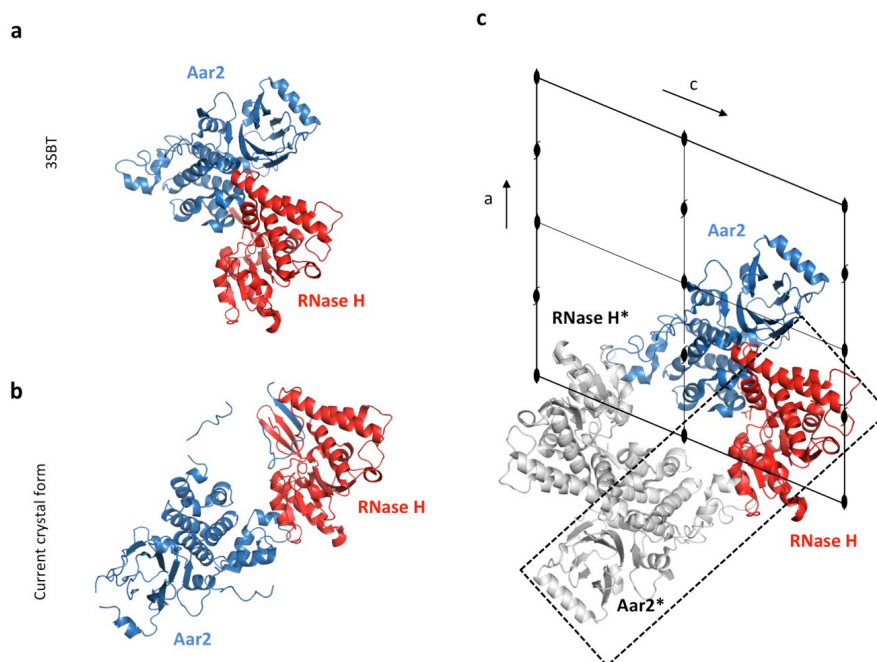
**Figure S11 | Interaction between Aar2 and the large domain of Prp8 (RT/Ln/En).** **a**, Aar2 binds to the junction between the Linker and Endonuclease domains burying 1230Å<sup>2</sup> of solvent accessible surface; **b**, specific interaction of arginine-928 located at the tip of the RT fingers domain with the helical region of Aar2; **c-g**, specific polar contacts between Aar2 and the Prp8 linker domain. Grey spheres represent water molecules.



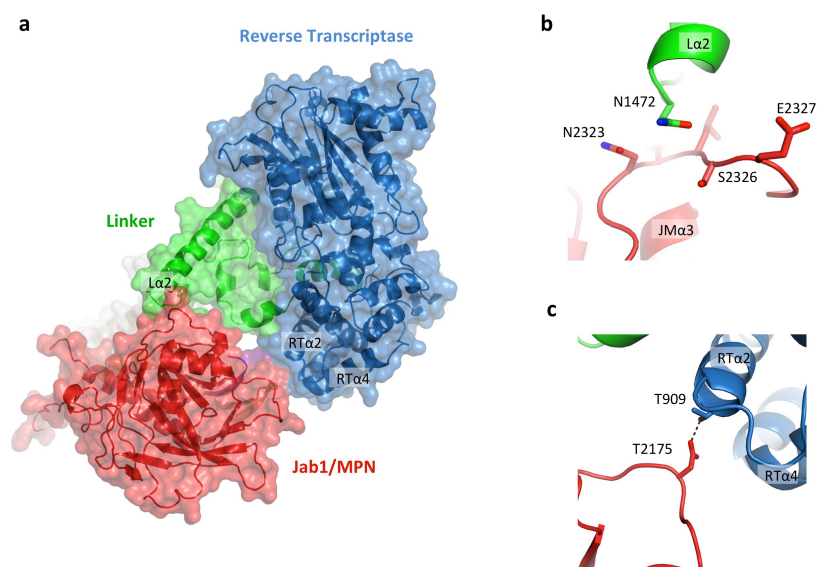
**Figure S12 | Molecular interface between the RNaseH-like and Jab1/MPN domains.** **a**, overview of the interacting surfaces; **b**, intermolecular parallel  $\beta$ -sheet between the Jab1/MPN domain, the C-terminal tail of Aar2 and the  $\beta$ -finger of RNaseH-like domain; **c,d**, specific recognition of the tail of Aar2 by Jab1/MPN and RNaseH-like domains; **e**, polar contact between helices RH $\alpha$ 5 and JM3<sub>10.1</sub>.



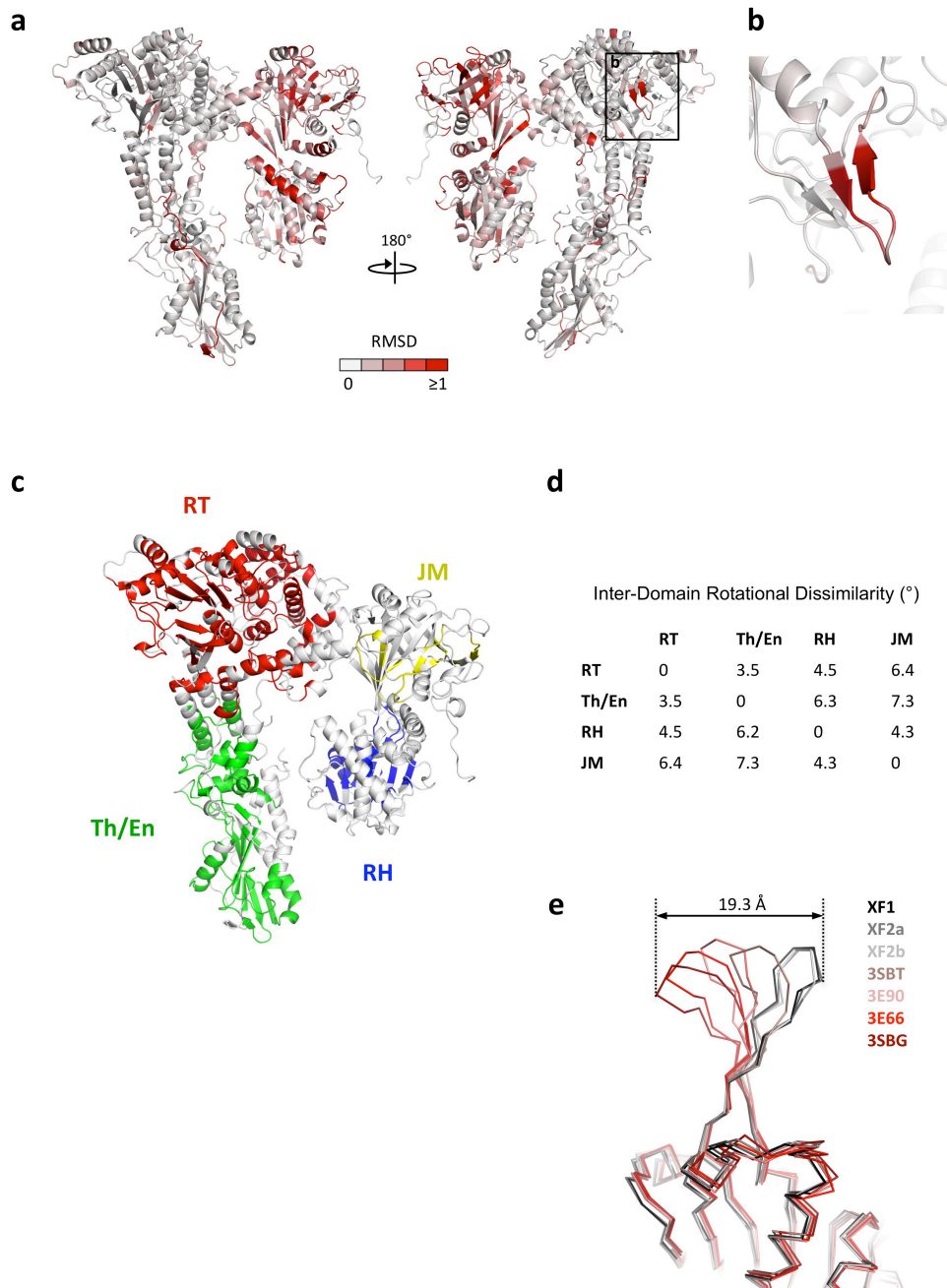
**Figure S13 | Molecular interface between Aar2 and the RNaseH-like domain.** **a**, overview; **b**, detailed interaction of the hydrophobic residues at the interface.



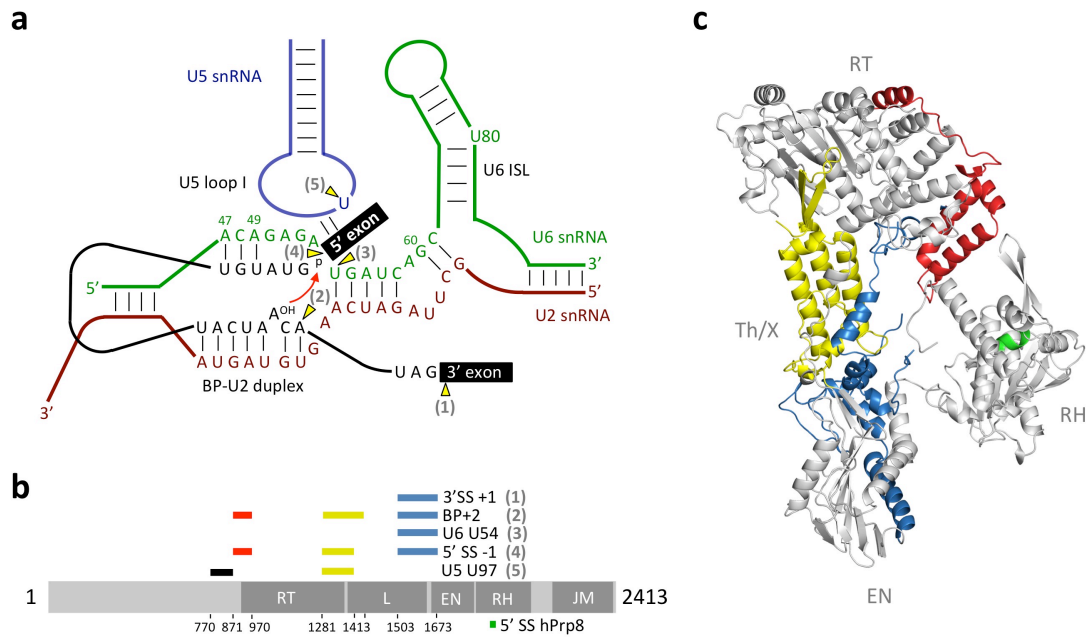
**Figure S14 | Interaction between Aar2 and the RNaseH-like domain of Prp8.** **a**, Molecular interaction in structure (PDB: 3SBT) described previously by Weber *et al.*<sup>29</sup>; **b**, Aar2-Prp8 interaction in the current crystal forms (C222<sub>1</sub> and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>). **c**, crystal contacts in the 3SBT crystal (C2 space group). Molecules related by a crystallographic 2-fold axis are shown in grey and marked with stars (\*). Dashed line indicates pair of molecules forming a contact virtually identical to those observed in both our crystal structures.



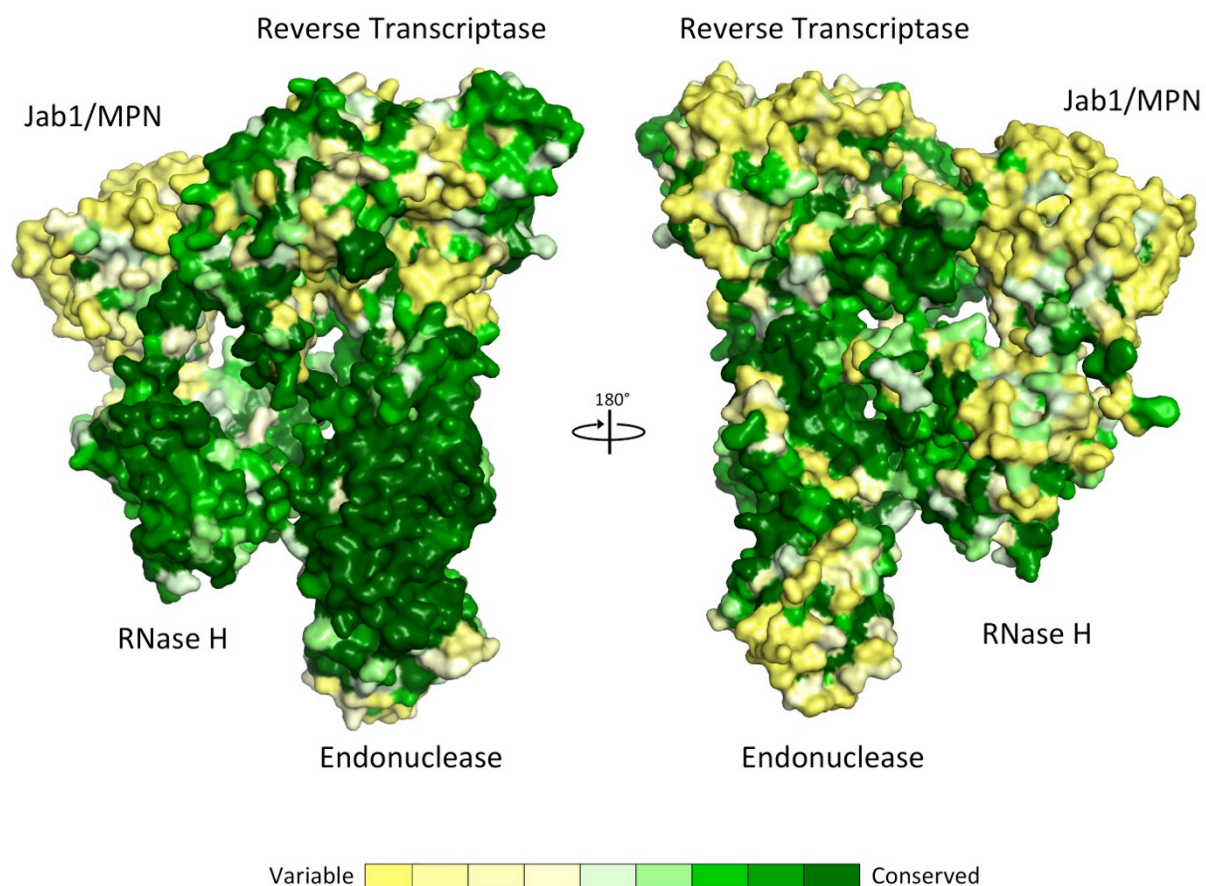
**Figure S15 | Interaction of the Jab1/MPN domain with the Linker and RT domains.** **a**, overview; **b** and **c**, contacts between the loops of the Jab1/MPN domain and the reverse transcriptase and linker domains.



**Figure S16 | Comparison of the Prp8 structure from P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> and C222<sub>1</sub> space groups using proSMART<sup>71</sup>.** **a,b**, local dissimilarity plotted onto cartoon representation of Prp8. **b**, highly variable β-finger (RTβ5-RTβ6) in RT domain; **c**, definition of the locally invariant rigid domains : RT - Reverse Transcriptase, Th/En – Thumb/X and Endonuclease domains, RH – RNase H-like domain, JM – Jab1/MPN domain; **d**, a matrix of Inter-Domain Rotational Dissimilarity, indicating relative rotational movement of the domain pairs, when aligned using various locally invariant rigid domains (defined in panel c); **e**, Superposition of the Prp8 RNase H-like domain from 7 crystal structures solved up to date, showing high flexibility of the β-finger. XF1 and XF2 refer to the C222<sub>1</sub> and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> crystal forms.



**Figure S17** | Contacts between Prp8 and various RNA components of the spliceosome. **a**, RNA network formed by pre-mRNA substrate, U2, U5 and U6 snRNA before first step of splicing. Red arrow represents attack of the branch point adenosine at the 5'SS. 4-thiouridine photo crosslinker incorporated at a specific residue is marked with yellow arrows: (1) 3'SS +1; (2) BP +2; (3) U6 snRNA U54; (4) 5'SS -1; (5) U5 snRNA U97. **b**, mapping of the site-specific cross-links on Prp8 sequence (ref 20). Bars indicate the smallest proteolytic fragments to which cross-linking sites were mapped: red, 871-970; yellow, 1281-1413; blue, 1503-1673; green, the region (1966-1970) corresponding to the 5'SS cross-link in hPrp8 (ref 18); **c**, mapping of the cross-linked regions on the Prp8 structure: RT, Reverse Transcriptase; Th/X, RT Thumb/X; EN, Endonuclease; RH, RNaseH-like.



**Figure S18 | Conservation of the surface residues in Prp8<sup>885-2413</sup> calculated with ConSurf server<sup>66</sup>.** Highly conserved residues map onto the surface of RNase H-like and Linker/En domains, as well as inside the active site cavity. Notably, a highly conserved patch of the Linker/En domain is also positively charged (Fig 5c), which is consistent with a role as the binding site for the RNA catalytic core (Fig 5b). In the Prp8-Aar2 complex Aar2 covers a large part of this surface.

**Table S1.** Data collection, phasing and refinement statistics

	Native1	Native2	Hg-2
<b>Data collection</b>			
Space group	C222 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	118.9, 158.6, 220.5	125.4, 177.8, 216.6	125.9, 178.2, 218.7
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	2.00 (2.03-2.00)	3.15 (3.15-3.10)	3.70 (3.81-3.70)
<i>R</i> <sub>merge</sub>	0.06 (1.26)	0.16 (1.29)	0.15 (1.91)
<i>R</i> <sub>pim</sub>	0.04 (0.62)	0.08 (0.59)	
<i>I</i> / $\sigma$ <i>I</i>	11.9 (1.5)	6.4 (1.4)	13.8 (2.4)
<i>CC</i> <sub>1/2</sub>	1.00 (0.57)	0.99 (0.63)	1.00 (0.76)
Completeness (%)	100 (100.0)	99.9 (100.0)	100 (100)
Redundancy	5.0 (5.0)	5.0 (5.2)	14.7 (15.6)
<b>Refinement</b>			
Resolution (Å)	48.3-2.0	137.4-3.1	
No. reflections	140 375	84 013	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	19.8/24.9	22.3/26.7	
No. atoms	14 902	28 284	
Protein	14 232	28 284	
Ligand/ion	0	0	
Water	670	0	
B-factors			
Protein	52.06	96.91	
Ligand/ion	NA	NA	
Water	47.37	NA	
R.m.s deviations			
Bond lengths (Å)	0.006	0.011	
Bond angles (°)	0.94	1.53	

**Table S2.**  
**Superposition of the domains in the current model (C222<sub>1</sub>) with the previously solved structures of Prp8 subdomains**

Domain	PDB code	Context of the structure	RMSD [Å]	Aligned region
RNaseH	3E9O	RNaseH alone	1.7	1832-2068
	3E66	RNaseH alone	2.4	1832-2068
	3SBT	RNaseH-Aar2 complex	1.3	1832-2068
	3SBG	RNaseH-Jab1/MPN	2.6	1832-2068
Jab1/MPN	2OG4	Jab1/MPN alone	1.3	2152-2394
	3SBG	RNaseH-Jab1/MPN	1.2	2152-2394
Aar2	3SBT	RNaseH-Aar2 complex	1.0	2-95, 99-153, 171-320
	3SBS	Aar2 alone	1.3	2-95, 99-153, 171-320

**Table S3.**  
**The residues (885-1375) of Prp8 show a polymerase fold**

PDB ID	Molecule	Chain:residues	Z-score	Aligned Residues	RMSD	Sequence identity
3SKE	HCV NS5B RdRP	A:1-517	13.9	310	3.8	7
3DU5	TERT	A:25-580	7.0	230	13.5	11
3OL8	Polio virus RdRP	A:1-461	11.9	237	4.0	10
3NAH	Norovirus RdRP	A:61-492	11.5	313	4.2	9
3HVT	HIV-1 RT	A:1-425	3.0	167	8.2	10

Structural similarity between the Prp8 (885-2413)-Aar2 complex and other proteins in PDB databases was assessed by secondary structure matching as implemented in PDBeFold web server (ref 63). Pairwise superpositions were made using DaliLite server (ref 64)

**Table S4.**  
**The residues (1650-1810) Prp8 have a type II restriction endonuclease fold**

PDB ID	Molecule	Chain:residues	Z-score	Aligned Residues	RMSD	Sequence identity
3HW4	PA endonuclease	A:1-198	6.1	98	3.3	9
1RVA	Eco RV	A:2-245	4.1	97	3.6	14

See Table S3 legend

**Table S5. Suppressors of 5'SS, 3'SS and BP mutations<sup>a</sup>**

<b>mutation</b>	<b>splice site mutations</b>	<b>location</b>	<b>references</b>	<b>remarks<sup>b</sup></b>
P986T	U6-U57 C or G and BP-A2G	Tip of helix near the external $\beta$ -sheet	35, 71	2nd step allele
M1399I	5'-SS, 3'-SS	Loop interior	36	
E1450K	3'-SS	pointing to the active site cavity	72	
L1557F	5'-SS, 3'-SS, BP	inner surface active site cavity	71	1st step allele
K1563I	3'-SS	inner surface active site cavity	73	
W1575R	5'-SS, 3'-SS, BP, U6 A51	Active site loop	71	2nd step allele
E1576V	5'-SS, 3'-SS, BP	Active site loop	71	2nd step allele
R1753K	BP	outer surface En domain	71	1st step allele
E1817G	5'-SS	Inner surface En domain	74	
L1823P	5'-SS, 3'-SS, BP	Loop leading to RNase H	35,71	2nd step allele
E1834L/S	Poly-pyrimidine binding	Loop leading to RNaseH	36	
T1861P	5'-SS and 3'-SS and U4-cs1	$\beta$ -finger RNase H	74, 37	
K1864E	5'-SS	$\beta$ -finger RNaseH	74	
N1869D	5'-SS, 3'-SS, BP	$\beta$ -finger RNase H	71,74,75,	2nd step allele
V1870N	5'-SS, 3'-SS, BP or U6-U57	$\beta$ -finger RNase H	35, 71	2nd step allele
E1960K/G	Poly-pyrimidine binding	Inner loop RNaseH	36, 71	1st step allele
T1982A	5'-SS, 3'-SS, BP-A2G	External loop	35,75	2nd step allele

<sup>a</sup>only single mutants are included

<sup>b</sup>First step alleles shown to improve the first step and inhibit the second. Second step alleles shown to improve the second step and inhibit the first (ref 71).

Table S6. U4-cs1 suppressors			
region	mutations	domains	locations
region a	R236G	unknown	unknown
	L261P		
	L280P		
region b	K611R	unknown	unknown
	E624G		
	N643S		
	V644A		
	D651G or N		
	H659P		
region c	K684E	unknown	unknown
	E788G or V		
	N796S		
	W856R		
	E860K		
region d	Q861R	RT domain	loop in 4 stranded $\beta$ -sheet loop in 4 stranded $\beta$ -sheet near in 4 stranded $\beta$ -sheet near in 4 stranded $\beta$ -sheet near in 4 stranded $\beta$ -sheet near in 4 stranded $\beta$ -sheet within loop following RT $\alpha$ 12 within loop following RT $\alpha$ 12 within loop following RT $\alpha$ 12
	D1094A or N or V		
	M1095T		
	V1098D		
	N1099K		
	I1104M		
	R1105L		
	P1191L or S or T		
	D1192Y		
region e	N1194D	endonuclease	top surface top surface top surface top surface top surface side surface side surface
	L1624M		
	L1634F		
	L1641F		
	T1685I		
	P1688L or R		
	A1754V	RNaseH	on inner surface $\beta$ -finger $\beta$ -finger $\beta$ -finger $\beta$ -finger
	N1809D		
	F1851L		
	V1860D or N		
	T1861P		
	V1862A or D or Y		
	I1875T		

All mutants are described in Kuhn & Brow (ref. 37) and Kuhn *et al.* (ref. 76).

---

**Table S7. *Brr2-1* suppressors**

---

mutations	domains	locations
L1066P	RT domain	near four-stranded $\beta$ -sheet
N1067K	RT domain	near four-stranded $\beta$ -sheet
N1087H	RT domain	near four-stranded $\beta$ -sheet
V1098I	RT domain	near four-stranded $\beta$ -sheet
L1107T	RT domain	near four-stranded $\beta$ -sheet

---

These mutants were described by Kuhn et al. (ref 37). Screening was carried out in the region 1022–1214, which spans the RT palm domain.

---

### Supplementary references

71. Liu, L., Query, C. C. & Konarska, M. M. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat. Struct. Mol. Biol.* **14**, 519-526 (2007).
72. Dagher, S.F. and X.D. Fu, Evidence for a role of Sky1p-mediated phosphorylation in 3' splice site recognition involving both Prp8 and Prp17/Slu4. *RNA* **7**, 1284-1297 (2001).
73. Ben-Yehuda, S., *et al.* Extensive genetic interactions between PRP8 and PRP17/CDC40, two yeast genes involved in pre-mRNA splicing and cell cycle progression. *Genetics* **154**, 61-71 (2000).
74. Collins, C. A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev* **13**, 1970-1982 (1999).
75. Siatecka, M., Reyes, J. L. & Konarska, M. M. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983-1993 (1999).
76. Kuhn, A. N., Li, Z. & Brow, D. A. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65-75 (1999).